An Analysis of Virtual Channel Memory and Enhanced Memories Technologies

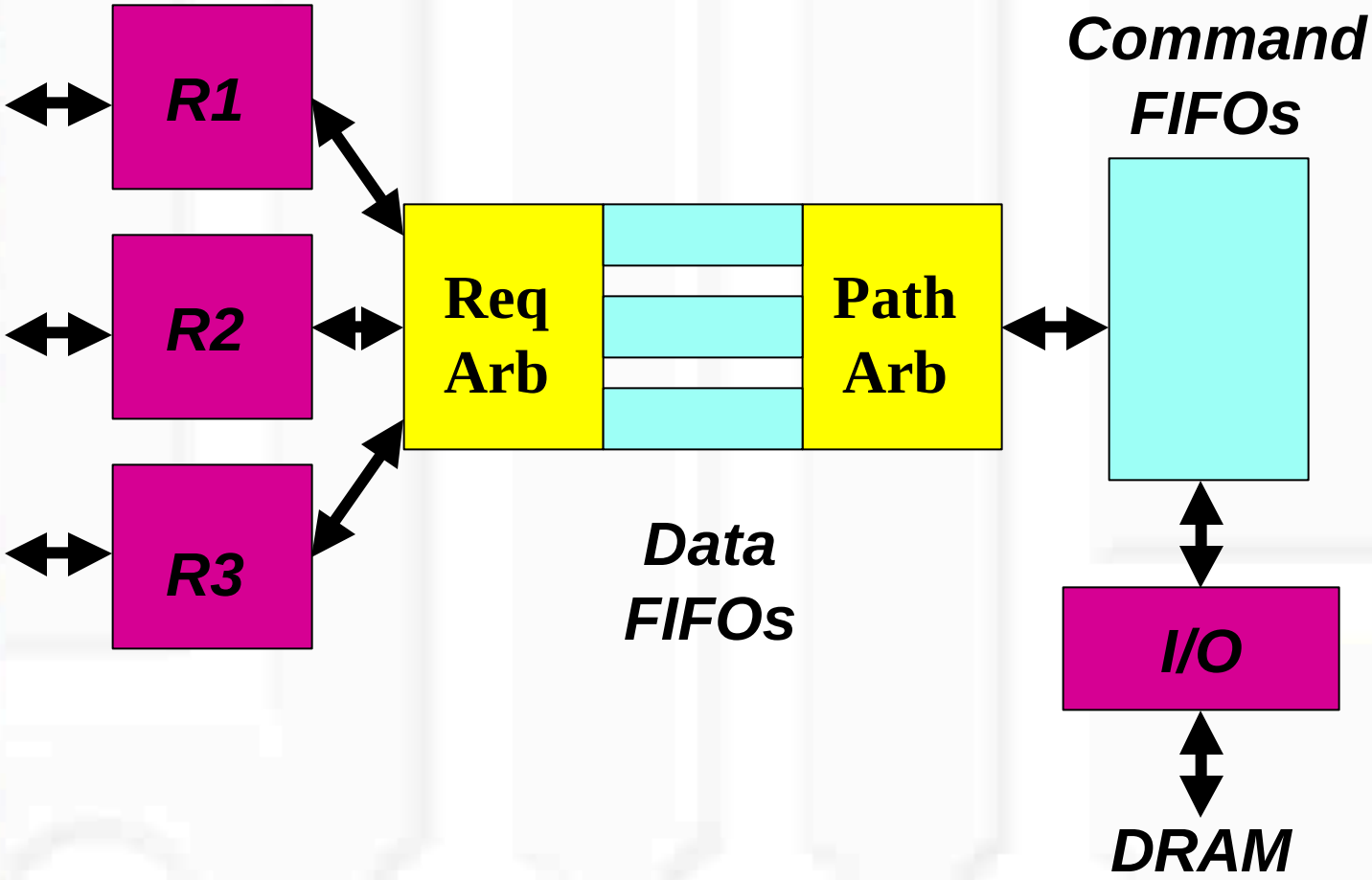
Bill Gervasi
Technology Analyst
Chairman, JEDEC Memory
Parametrics Committee

San Jose January 23-24, 2001 • Taipei February 14-15, 2001

Agenda

- Next Generation PC Controllers
- Concerns With Standard SDRAM
- Cached SDRAMs:
  - Enhanced & Virtual Channel
- Controller Complexity vs DRAM type
- Pros and Cons of Cached DRAMs
- Conclusions & Call to Action

Next Gen PC Controllers



Requestors

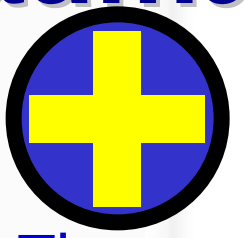
- CPU port: Cache fills dominate
  - DRAM frequency =

    1/3 to 1/5 of CPU frequency
  - Big L2 caches randomize memory accesses
- Graphics port: Lots of random accesses
  - Especially 3D rendering
- South Bus port: Mix of short, long packets

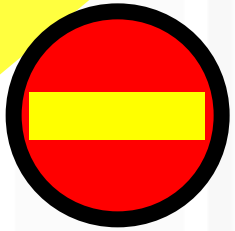
SDRAM Roadmap

- DDR II
- DDR-333
- DDR-266
- PC-133
- PC-100

SDRAM 66



The good news: ever faster cores, power manageable, simple evolutionary changes, at about the same price



The bad news: random access time has not changed appreciably, and power is higher than it needs to be

Concerns With SDRAM

1. Power
2. Latency
3. Refresh Overhead

SDRAM Power vs Latency

- Active power is very high
  - Active on power 500X inactive off power
  - Encourages controllers to close pages
- But access time to a closed page is long
  - Row activation time + column read time

SDRAM Power Profile

Open Page

Closed Page

Power State*	Relative Power	CPU Clock Latency**
Active on	100%	0 x 5 = 0
Inactive on	12%	3 x 5 = 15
Active off	4%	1 x 5 = 5
Inactive off	0.2%	4 x 5 = 20
Sleep	0.4%	200 x 5=1000

* Not industry standard terms – simplified for brevity
** Assuming memory clock frequency = 1/5 CPU frequency

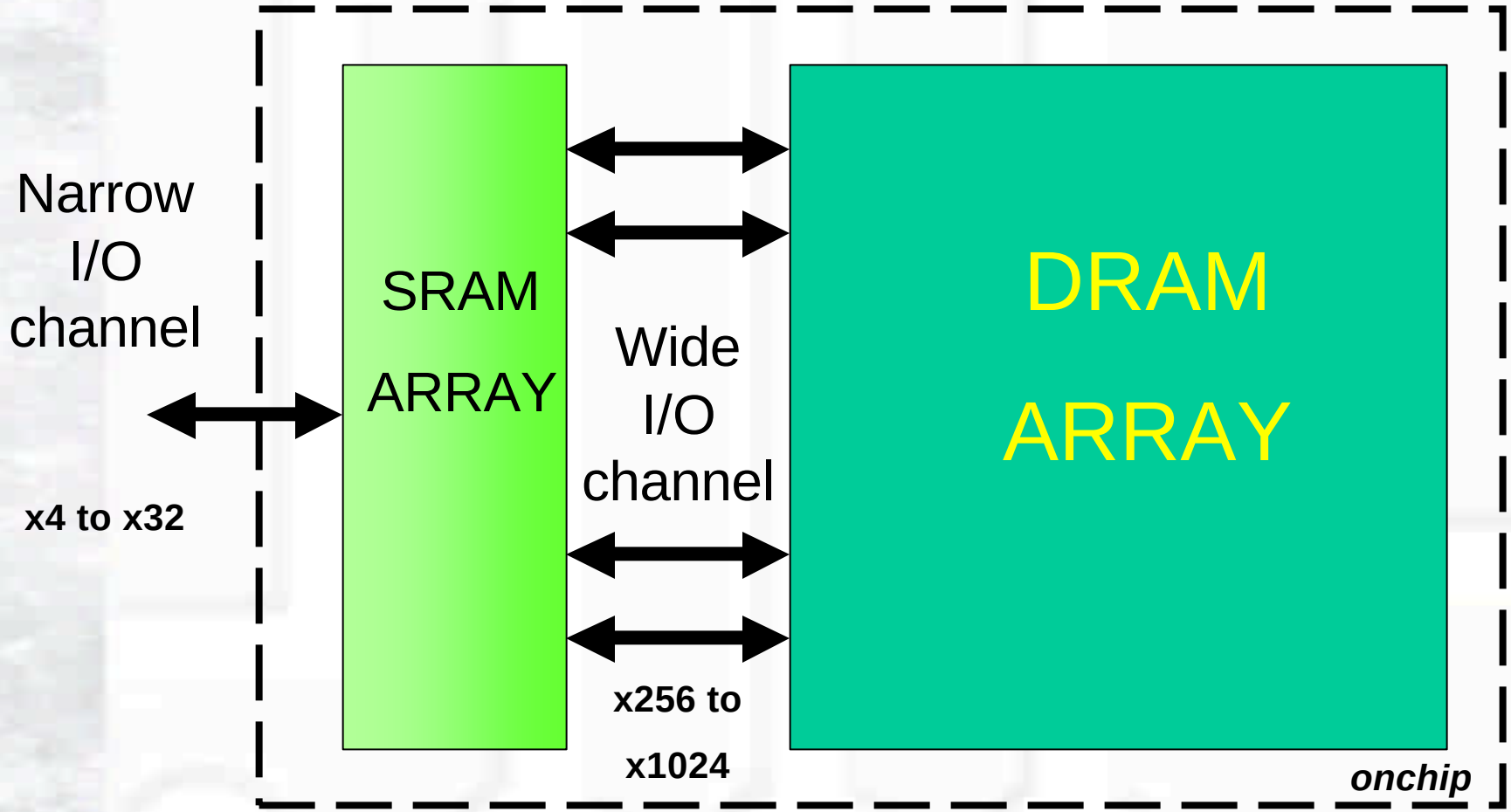
Refresh

- Gigabit generation refresh overhead
  - 256Mb generation is 75ns each 15.6us
  - 1Gb generation will be 120ns each 7.8us
- ***This is a 3X performance penalty***

An Argument for Cached SDRAM Architectures

Platform Conference

What Are Cached SDRAMs?



Cached DRAM architectures can address SDRAM's key limitations

However, only commodity memories are affordable for mass market use

I hope to encourage the adoption of cache for all standard future SDRAMs

Cached SDRAM Solutions

1. **Power:** Encourage closed page use
2. **Latency:** Fast access to closed pages
3. **Refresh:** Background operation

Cached DRAM

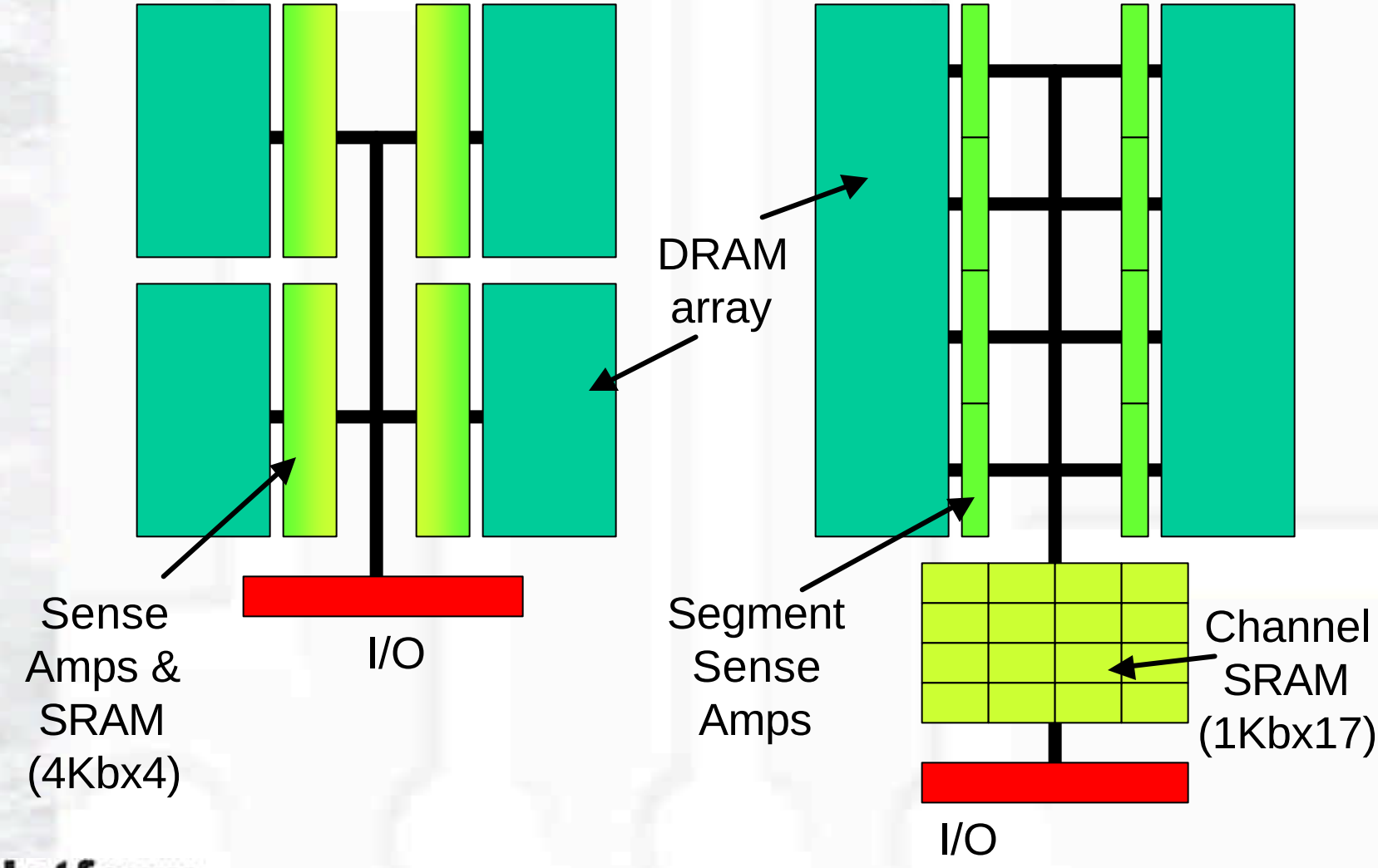
- SRAM cache before DRAM array
  - Much like CPU onchip caches
  - Exploit wide internal buses for fast block transfer of DRAM to SRAM and back
  - Allow DRAM core to be deactivated while…
  - … I/O performed on SRAM

Two Leading Cached DRAMs

- **Enhanced SDRAM**
  - SRAM in sense amps
  - Direct mapped into array
- **Virtual Channel SDRAM**
  - SRAM in periphery
  - DRAM associativity maintained by controller

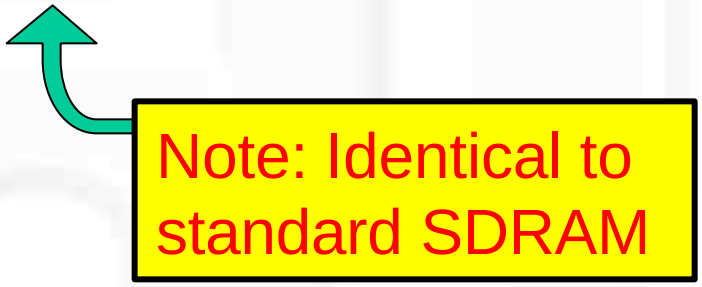
Note: Other cached DRAM architectures exist, however none have been proposed as a commodity DRAM standard.

Enhanced & Virtual Channel



Basic Functions

- Enhanced
  - Activate
  - Read (& cache fill)
  - Write (array & cache update)
  - Precharge
  - Auto Refresh
  - Self Refresh
- Virtual Channel
  - Activate
  - Prefetch (cache fill)
  - Read (from cache)
  - Write (cache update)
  - Restore (array update)
  - Precharge
  - Auto Refresh
  - Self Refresh



Power Factors

- Enhanced
  - Page open to read to cache
  - Closed while reading
  - Open during writes
- Virtual Channel
  - Page open to read to cache
  - Closed while reading
  - Closed during writes
  - Reopen for restore

Latency Factors

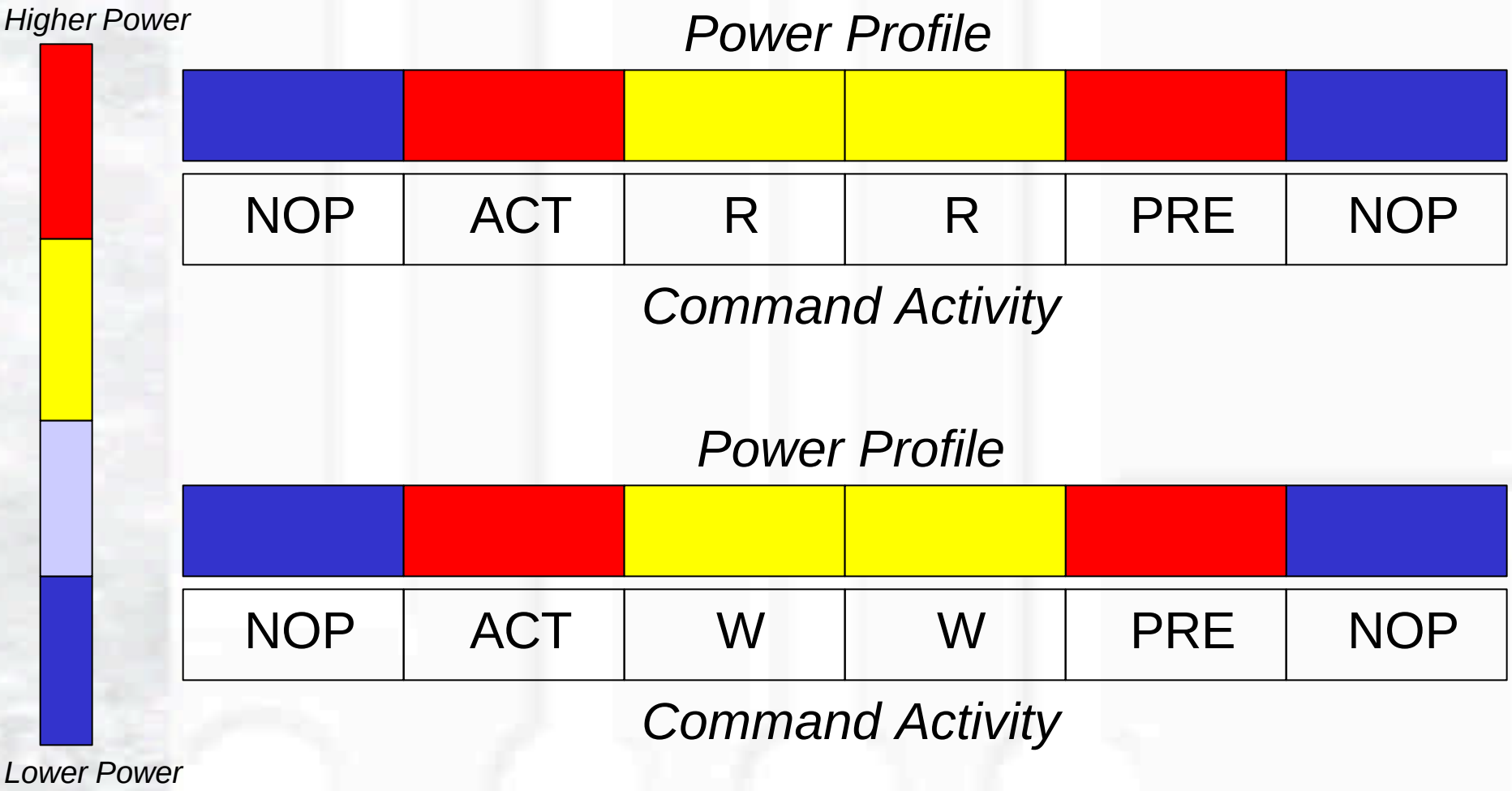
- Enhanced
  - Read automatically fills cache
  - Reads on closed page permitted
  - No write latencies
  - Masking increases write recovery time (normal for SDRAM architectures)
- Virtual Channel
  - Prefetch for cache fill
  - Reads on closed page permitted
  - Writes on closed page permitted
  - Reactivation to restore
  - Inherently RMW – no penalty for masking

Refresh Factors

- Enhanced
  - Activation not allowed
  - Permits reads during auto refresh
- Virtual Channel
  - Activation not allowed
  - No reads during auto refresh

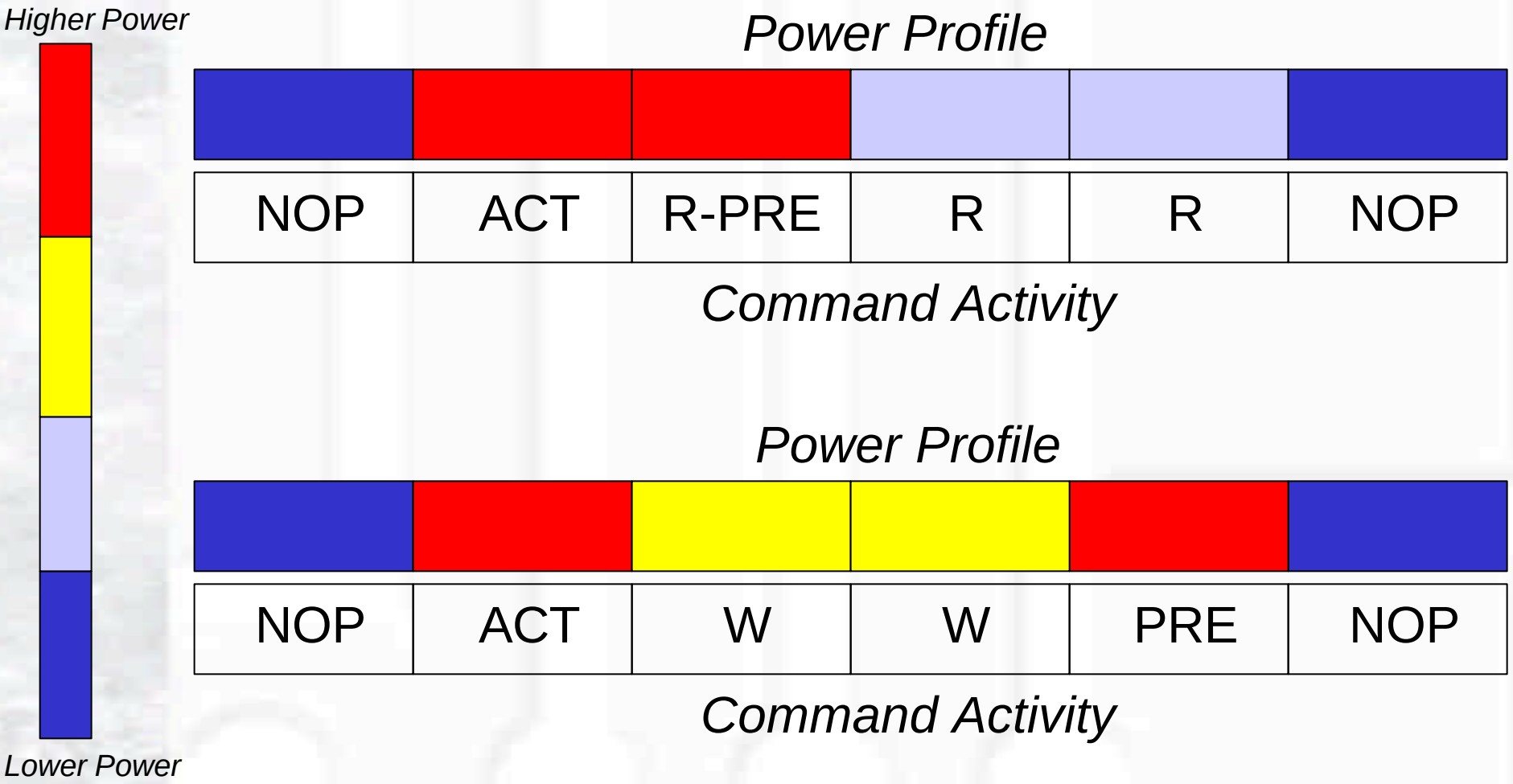
Let's Compare the Power and Activity Profiles of SDRAM, Enhanced, and Virtual Channel

Closed Page SDRAM Profile



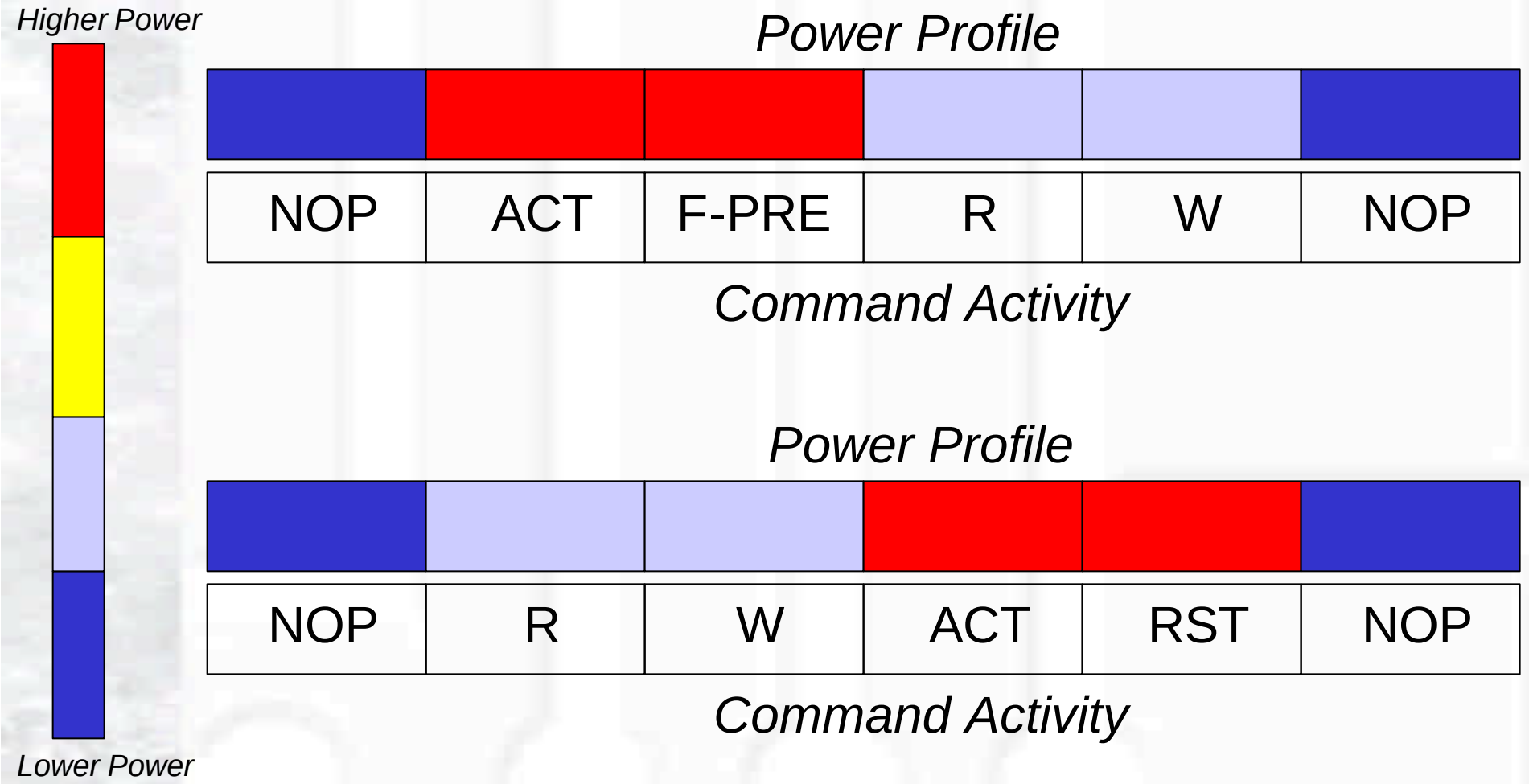
Platform Conference

Enhanced Profile



Platform Conference

Virtual Channel Profile



Platform Conference

Power Profile Factors

- Standard L3 cache hit analysis
- Profile of memory operations affected by randomness of accesses
  - Balance of activations & precharges, reads & writes
  - Requestor channel profile depends on application – games, video, or office app?

Cache Entry Size

- Max words burst per hit before reload needed

  Entry size ? bus width ? burst length

Burst length = 4	x4	x8	x16
Enhanced	**256**	**128**	**64**
Virtual	**64**	**32**	**16**

- Miss impacts performance & power
- Affects controller association overhead
- Die size impacted by entry size

Randomness

- Enhanced controllers replace any entry
  - Physical memory locality determines
- Virtual controllers can lock channels, e.g.
  - Screen refresh channel never replaced
  - Priorities can be assigned to channels
  - Weighted replacement algorithms possible

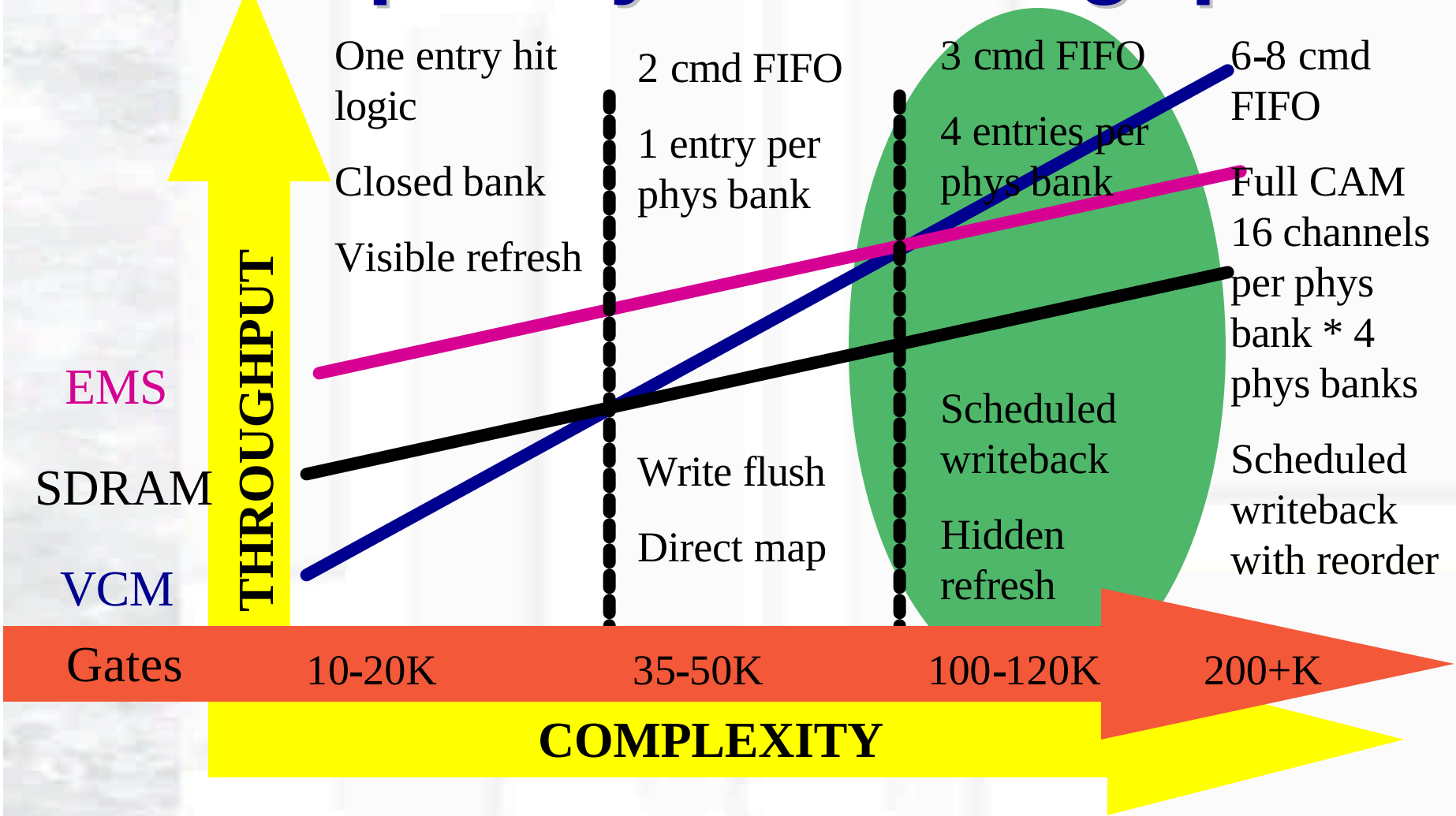
Enhanced Controller Issues

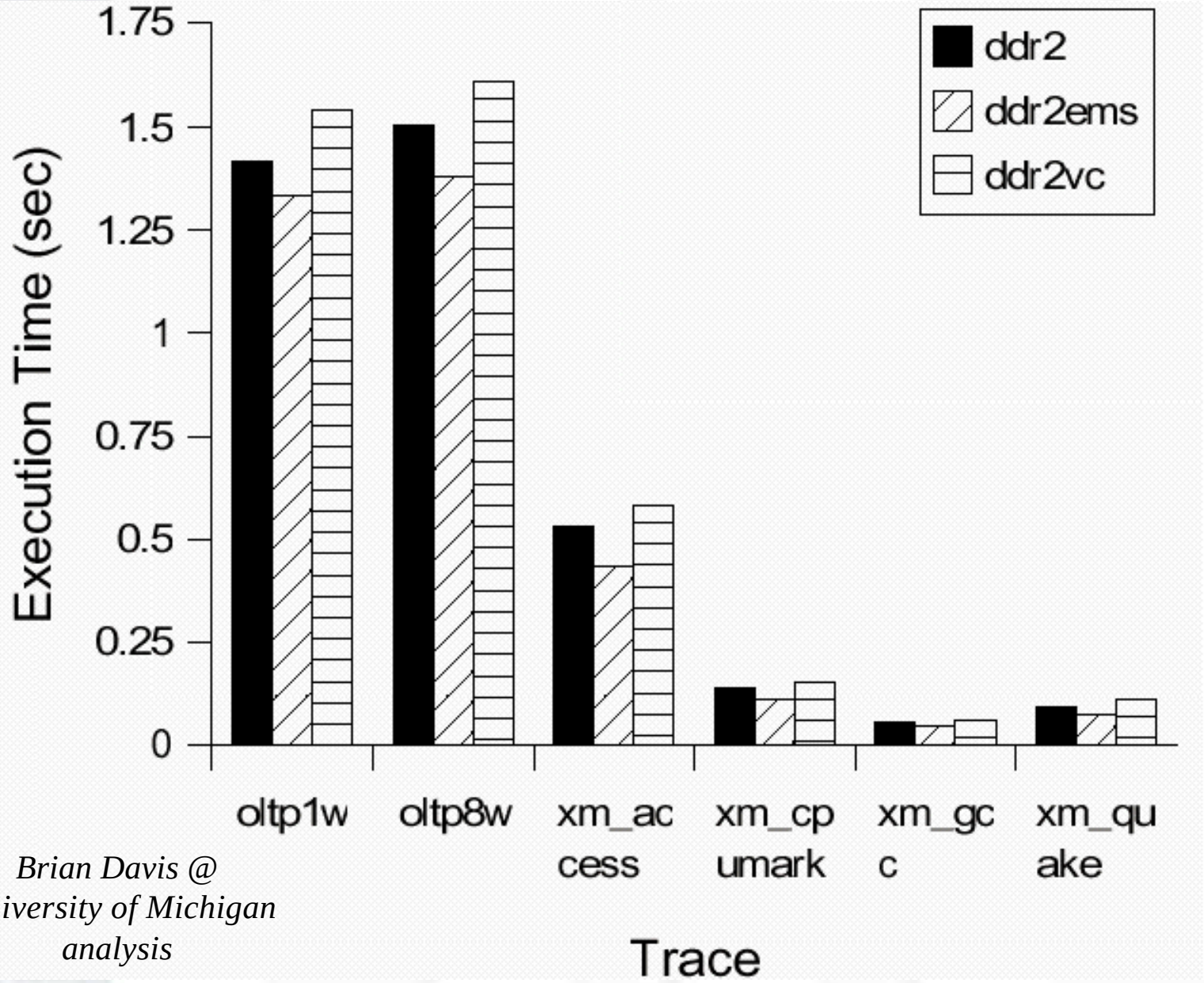
- Direct mapping restricts complexity
- 4 physical banks x 4 cache lines
- Comparitor for cache hit

Virtual Controller Issues

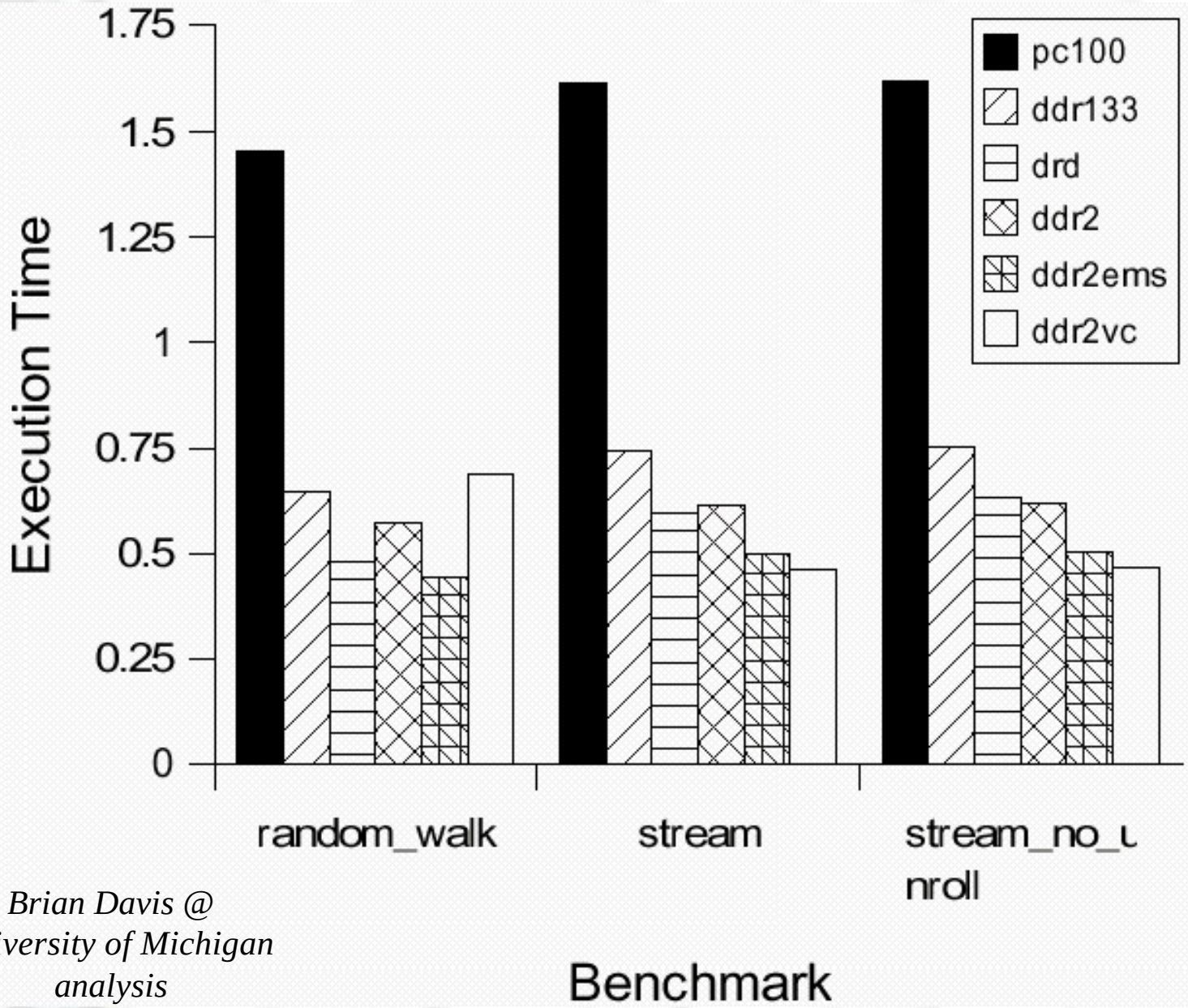
- Tags required to do closed page writes
  - Keep track of where to restore
- 4 physical banks x 16 channels
- CAM needed to exploit large # channels
  - CAM routing challenging – metal over array?

Complexity vs Throughput





Brian Davis @ University of Michigan analysis



Brian Davis @ University of Michigan analysis

Virtual Channel

- PROS
  - Best performance headroom from cache associativity
  - Flexible channel maps
  - Saves power on reads and writes
  - More hits in random access environment
  - Pin compatible in SDR & DDR I configurations
- CONS
  - Die penalty from 7-13%
  - Incompatible command set
    - **Cannot have one die for standard and virtual**
  - Refresh blocks access
  - Simple controller performance < “standard”
  - Small cache entry sizes cause more replacement

Enhanced Memory

- PROS
  - Die penalty from 2-5%
  - Compatible command set
    - **Same die supports standard or enhanced**
  - Optional features
  - Simple to implement
  - Pin compatible
  - No performance drop even for simple controller
  - Refresh does not block reads
- CONS
  - Royalty to DRAM suppliers
  - Performance boost lower than VCM max
  - Less flexible maps
  - No power savings on writes

User Perspective

- Users want a standard cached DRAM architecture – solves real system problems
  - Lower power for closed bank policy
  - Lower latency
  - Hidden refresh
- Costs cannot exceed 5% over non-cached

Conclusions

- Cached DRAM is highly desirable
  - Both Enhanced and Virtual offer significant benefits
- VCM requires more controller complexity
  - Low end not well served
- Next “commodity” must fit all markets
- Enhanced preferred over Virtual Channel
  - Better chance at becoming “standard”
  - Compatibility, simple transition lowers adoption risk
  - Like L2 went direct ✍ set associative, maybe VCM is a good fit for DDR III in 2006/2007

Call to Action

- Act now!
- Demand vendors provide cached DRAM
- Express support for Enhanced style
- Commit controller designs

Industry likely to decide in March 2001

Platform Conference

Platform Conference